

NMI Performance Inside, Grant #0438372
Annual report
Jennifer Schopf, July 27, 2005

The NMI Middleware Development and Prototyping project “Performance Inside” has concentrated on developing software and tools to allow users to better understand the basic behavior of Grid systems. The goal of these tools is to eventually allow users to understand the baseline performance behavior of distributed systems and applications so that the changes that occur can be easily understood and fault behavior diagnosed. The tools will support the development of extended monitoring frameworks that enable the capture of performance data across the environment by using a set of sensors, or probes, widely deployed to gather basic data.

In the first year of the project (start date: Oct 2004), we have concentrated on solidifying a basic infrastructure and making sure that it is scalable and dependable, extending the basic set of sensors, or information providers, and developing a Trigger service to provide early fault detection. We have also identified early users and begun work with specific user communities.

We have used the NMI-packaged Globus Toolkit Monitoring and Discovery System, MDS4, as our basic starting point. MDS4 provides basic sensor mechanisms, several information provider implementations, logic for querying and subscribing to information providers, and a simple archiver, all integrated with the NMI PKI security infrastructure. In addition, this system is fully compliant with current web service standards, thereby providing standard query and subscription interfaces, as well as seamlessly integrating into any WSRF compliant information source.

We have implemented a Trigger service that monitors the conditions of other services and executes a secure, administrator-defined executable programs when user-defined conditions are met. The Trigger service is built using the basic MDS4 infrastructure; it uses XPath as condition-matching language, and periodically executes a user-defined XPath expression at user-defined intervals. Users or administrators can plug-in different types of action scripts that can be executed when different conditions are satisfied. We have also implemented a rate-limiting mechanism that throttles the execution of the action scripts when conditions stay in a steady state for extended (user-defined) periods of time.

The a prototype of the Trigger service is currently in active use by the Earth Science Grid project, and they are in the process of deploying the GT4 version. It will gather data from a set of services (RLS, SRM, OpenLDAP, the HTTP file server, and the GridFTP server), and send email to system administrators when these services go down. We are continuing work on this collaboration especially to make deployment easier.

We have been extending the information sources of MDS4. This has begun with interfaces to Hawkeye and Ganglia, for cluster status information, and PBS and LSF, for basic queue information. As part of our work with the TeraGrid project we are extending

the set of attributes we currently gather from these sources, and adding additional interfaces to CluMon, a cluster monitoring system developed at NCSA, and the Torque Resource Manager, a derivative queuing system related to OpenPBS. These extensions will be used by the various TeraGrid Metascheduling approaches, as well as a user-based portal to make resource selection decisions. We have decided to wait on interfacing to Inca until their next revision (currently underway) is completed.

Under separate funding, we have also extended WebMDS, an extensible web-browser based display and query tool for WSRF services. WebMDS is implemented as Java servlet and uses XSLT transforms on the data. A sample page is located at <http://mds.globus.org:8080/webmds/webmds?info=indexinfo&xsl=servicegroupxsl>

In terms of research, we have begun to perform an in depth performance analysis of the MDS4 index server, and submitted a paper to SC 2005 (<http://www-unix.mcs.anl.gov/~schopf/Pubs/mds4.sc.pdf>). Basic results show that for small indexes we achieve very good behavior (~76 requests/sec) but as the index size grows the performance decreases significantly. We are currently investigating the bottlenecks in the system to better understand if this is an implementation issue, or a standards-definition issue associated with WS-N. We have seen excellent results in terms of reliability of the index server. We have had a long-running test ongoing for over 80 days (as of July 25) that has handled over 500 Million queries with virtually no load on the system.

We are currently working with two user communities, ESG and TeraGrid, and plan to begin work with the Open Science Grid (OSG) project shortly. For ESG we are extending the Trigger service for failure notification. For TeraGrid we are working on providing a unifying interface to information sources for use in resource selection decisions, specifically for use by one of the TeraGrid Metascheduling systems being designed but also by users through a portal.

We have also improved the basic infrastructure of MDS4 as a part of the Performance Inside grant work to make it more stable and accessible to users. In the 4.0.1 release, due late July 2005, MDS4 will be tomcat-compliant so will be able to run in a tomcat container as well as the standard Globus Toolkit container. We have significantly improved the security of MDS4 and held a meeting with several NMI security experts (V. Welch, J. Basney, R. Ananthakrishnan) in early July. The documentation and testing of all MDS4 components has also been significantly upgraded.

Some additional information is available from <http://www.mcs.anl.gov/~jms/PerfInside/>